

The Availability, Quality, and Use of Administrative Data to Identify and Reduce Health and Human Services-Related Disparities: A Survey of HHS-Funded Programs and Interventions

KEY POINTS

- Executive Orders No. 13985 (2021) and 14091 (2023) require federal agencies to ensure that their policies and services, including data collection, advance equitable outcomes for all populations.
- The Assistant Secretary for Planning and Evaluation (ASPE) and the U.S. Department of Health and Human Services (HHS) Data Council Equity Data Subcommittee sought to learn more about how administrative data on demographic characteristics is used to assess service equity in HHS-funded programs.
- In the summer of 2022, RTI International, under direction from ASPE, designed and administered a web-based survey of HHS-funded programs delivering services to individuals.
- The survey identified wide variation in the types of demographic data collected and data collection methodologies used by HHS programs both within and between HHS offices and agencies.
- The quality of demographic data collected by HHS programs varied based on three key factors: the methods used to collect the data, what questions were asked, and which response options were offered.
- The results from the survey also indicated three opportunities to make data collection more robust and usable for future equity analyses. These include: 1) Increase the use of data collected through participant self-report 2) Increase participant response rates 3) Standardize response options across programs.
- In considering changes to administrative data collection, program staff should be included in decision-making processes to confirm that new data collection domains and response options align with program foci and will be useful for programmatic decision-making.

I. BACKGROUND

Executive Orders [13985 \(2021\)](#) and [14091 \(2023\)](#) require federal agencies to ensure that their policies and services, including data collection, advance equitable outcomes for all populations. The Executive Orders specifically emphasize the importance of ensuring equitable service provision for “...individuals who belong to communities that often have been denied such treatment, such as Black, Latino, Indigenous and Native American, Asian American, Native Hawaiian, and Pacific Islander persons and other persons of color; members of religious minorities; women and girls; LGBTQI+ persons; persons with disabilities; persons who live in rural areas; persons who live in United States Territories; persons otherwise adversely affected by persistent poverty or inequality; and individuals who belong to multiple such communities.”¹

¹ Exec. Order No. 14091, 88 Fed. Reg. 10825 (February 16, 2023).

Administrative data, as defined by OMB Memo M-14-06 are "...data held by agencies and offices of the government or their contractors or grantees (including States or other units of government) and collected for other than statistical purposes."² Administrative data are typically collected and used for programmatic and regulatory purposes such as basic program administration and tracking the provision of services. Agencies and offices across the Department of Health and Human Services (HHS) collect administrative data³, often in the form of demographic data, from program beneficiaries as a routine part of service delivery. An HHS Data Council working group sought to better understand the types of administrative data collected by HHS and how those data are used to examine service equity within HHS programs. This research effort will inform the development of future data policy ensuring that HHS collects administrative data in ways that facilitate its use to support ongoing program improvement and equitable service delivery.⁴

What is the HHS Data Council?

The HHS Data Council (Council) is the principal internal advisory body to the HHS Secretary on health and human services data policy. Consisting of senior level officials and staff representatives from HHS operating and staff divisions, the Council is co-chaired by staff in the Office of the Assistant Secretary for Planning and Evaluation (ASPE) and the Director of the National Center for Health Statistics. The Council coordinates data policy activities within HHS and conducts research to improve the collection and use of HHS data. The Council also supports the Data Governance Board by addressing analytic and technical data policy issues and the Chief Data Officer is represented on the Data Council.

II. APPROACH

Under contract to ASPE, RTI International conducted a web-based survey of HHS staff with knowledge of demographic data from a sample of programs delivering services to individuals. As this was the first attempt to explore how programs across HHS collect and use their demographic data, there was no validated survey instrument available for use in this project. Therefore, the research team consulted several existing surveys to develop a novel web-based survey instrument.^{5,6,7}

² Statistical purposes, as defined by OMB [Memo M-14-06 Guidance for Providing and Using Administrative Data for Statistical Purposes](#), refers to "the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups," (PL-107347, Title V—Confidential Information Protection and Statistical Efficiency Act (CIPSEA), Section 502 (9)(A)). Statistical purposes exclude "any administrative, regulatory, law enforcement, adjudicatory, or other purpose that affects the rights, privileges, or benefits of a particular identifiable respondent" (PL-107-347, Title V—CIPSEA, Section 502 (5)(A))."

³ Administrative data, as defined by OMB [Memo M-14-06 Guidance for Providing and Using Administrative Data for Statistical Purposes](#), are "...data held by agencies and offices of the government or their contractors or grantees (including States or other units of government) and collected for other than statistical purposes. Administrative data are typically collected to carry out the basic administration of a program, such as processing benefit applications or tracking services received. These data relate to individuals, businesses, and other institutions"

⁴ [Per Statistical Policy Directive 15 \(SPD15\)](#), "A combined race and ethnicity question is required for both self-response and proxy data collection. Respondents shall be offered a single combined race and ethnicity question that allows them to select one category or multiple categories. A single selection will be considered a complete response (e.g., Hispanic or Latino respondents are not required to select an additional category). SPD 15 requires the collection of detailed data on race and ethnicity beyond the minimum categories, unless an agency determines that the potential benefit of the detailed data would not justify the additional burden to the agency and the public or the additional risk to privacy or confidentiality, and therefore requests an exemption from OIRA."

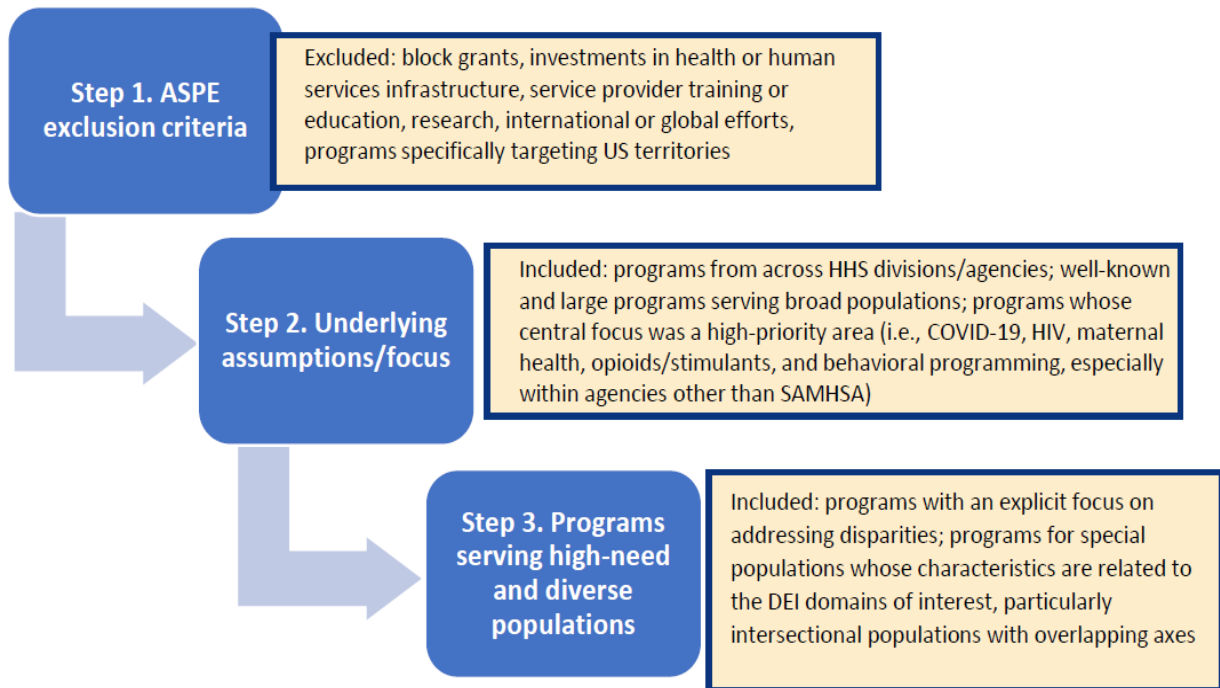
⁵ American Hospital Association (AHA). (2020). AHA Annual Survey. Retrieved from <https://www.ahadata.com/system/files/media/file/2022/05/2020-AHA-Annual.pdf>.

⁶ National Academies of Sciences, Engineering, and Medicine (NAEM). (2017). *Data management and governance practices*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/24777>.

⁷ Northwestern University. (2014). Report on Data Management Survey, Northwestern University. Retrieved from https://arch.library.northwestern.edu/concern/generic_works/xw42n7888.

The initial sampling frame included 472 HHS-funded programs listed in the [Assistance Listings Database](#)⁸. This frame was narrowed down through a three-step process ([Exhibit A-1](#)) to a final sample of 66 programs. The research team invited one key staff person, via email, from each of the 66 programs to participate in the web-based survey. Respondents replied over a period of three-and-a-half weeks during the summer of 2022. Of the 66 programs invited to participate, 29 programs (44%) submitted complete survey responses. An additional six programs (9%) submitted partial responses (i.e., responded to some questions, but discontinued the survey prior to responding to all questions). The partial responses are included in the counts for question response data when available.

Exhibit A-1. Sample Selection Process After Identifying 472 HHS-Funded Programs



III. FINDINGS

This study was guided by seven research questions ([Appendix A](#)) regarding the types of data collected, data collection processes, data use, and overall data quality. Additionally, the survey explored the extent to which the same demographic variables are used across HHS programs.

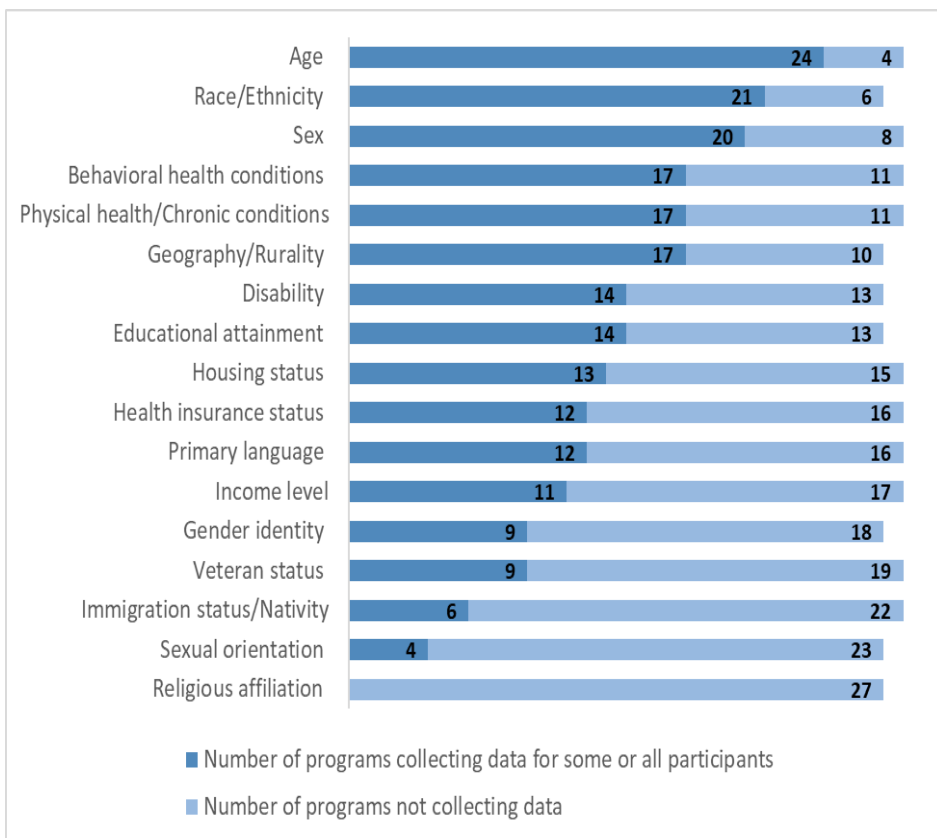
Research Question 1: What types of demographic data are HHS programs collecting about individuals enrolled in/served by HHS-funded programs?

Although all programs were collecting some demographic data, the breadth and depth of collection varies between programs ([Exhibit A-2](#)). Most programs collected data for age (83%), race/ethnicity (73%), and sex (69%). More than half of programs collected data related to behavioral health conditions (59%),

⁸ The Assistance Listing Database is part of the System for Award Management (SAM.gov). It is an official website of the U.S. Government that can be used to access publicly available award data via data extracts and system accounts.

physical health needs or chronic physical health conditions (59%) and geography/rurality (59%). Less than half of programs collected data related to disability or functional impairment status (49%); educational attainment (49%); housing/housing status (45%); health insurance status (41%); primary language (41%); income level/poverty status (38%); gender identity (31%); and veteran status (31%). Few programs collected immigration status/nativity (21%) or sexual orientation (14%). None of the programs surveyed collected data on religious affiliation. For each of these demographic characteristics, 3%-6% of survey respondents were unsure whether it was included in their administrative data collections.

Exhibit A-2. Data Collection Status by Demographic (n = 29)



Research Question 2: Is the demographic data collected by HHS of sufficient quality to assess disparities in service utilization, program outcomes, and program impact?

The quality of demographic data collected by HHS programs varied based on three key factors: the methods used to collect the data, what questions are asked, and which response options were offered. For example, among the 24 programs that collected age data, 54% collected it by birth date and 46% collected it by age span (e.g., under 18, 19-30, 31-50 years old, etc.). Respondents indicated that different question types yielded different levels of data detail and introduced limitations for how those age data could be used, analyzed, or shared by programs. For example, collecting an individual’s full birth date allows greater options for analysis as compared to collecting age via age span alone.

Most survey respondents, 83% (N= 24), also noted issues with missing data. While there may be many causes of missing data, providing respondents with an option to skip demographic questions is a source of “missingness” common across multiple programs in this sample. Seven (29%) of the twenty-four programs collecting age data allowed program participants to skip this question, potentially reducing the generalizability of future analyses performed with age data from those seven programs.

Another issue noted by respondents related to the methods used to collect demographic data. While self-identification is the approach preferred in most instances by the Office of Management and Budget

(OMB)⁹, 36% of respondents indicated that some or all their program's demographic data were collected through observation. This includes 14% of programs (3 of 21) that collected race/ethnicity data, 15% of programs (3 of 20) that collected sex data, and 22% of programs (2 of 9) that collected data about gender identity.

Research Question 3: To what extent are demographic data used to understand health and human-services related conditions, services, and outcomes of individuals and different communities?

Nearly all respondents (93%) indicated that they used demographic data to help their program track the total number of people or families served, volume of services received, and beneficiary demographic characteristics. Demographic data also were used to gauge contractor or grantee program success and overall program quality. As one respondent shared, "Internal analyses are used to assess awardee performance, support the identification of technical assistance needs, and identify priorities for continuous quality improvement initiatives." Another program highlighted data application to promote cross-program enhancements, explaining that data support their ability to "Share lessons learned and best practices for culturally responsive adaptation of evidence-based strategies for [the program]."

Access to demographic data was also credited with enabling greater awareness of unmet needs and supporting service expansion planning. One respondent said that program staff used analyses of program data to "advocate for housing access (shelter is the most commonly cited service need)." Another shared that data documented a "need for increased programming to rural populations," and a third explained that data were "a fiscal indicator for securing additional monies for urban programs." Two other respondents specifically noted that demographic data helped their programs understand service use during the COVID-19 Public Health Emergency, which underscored not only the applicability and usefulness of these data, but also the importance of having immediate access to current data to address emergent needs.

Research Question 4: What are the barriers to routinely collecting and using demographic data?

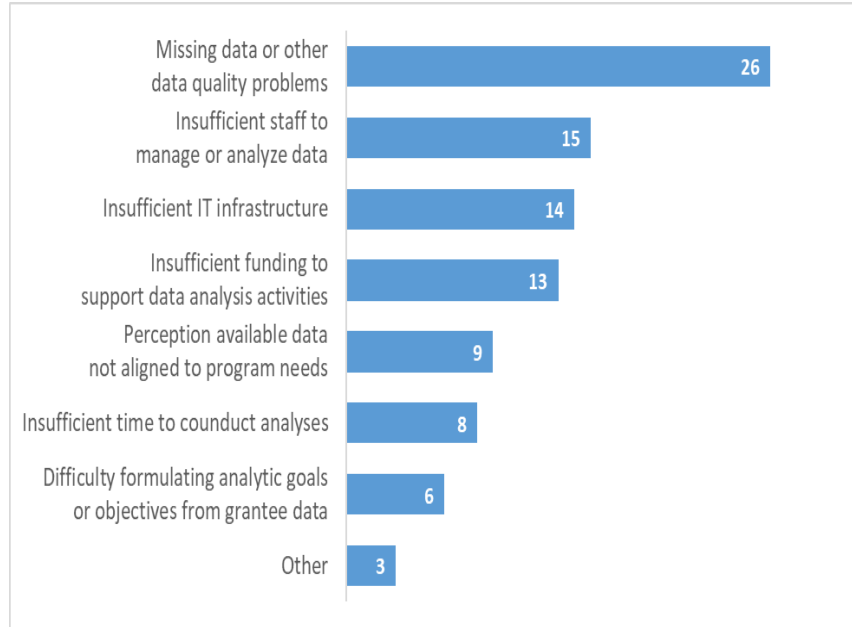
Some programs reported only collecting demographic data explicitly needed to administer services or to comply with federal guidelines. Other barriers reported to the collection of a wide range of demographic variables included perceived staff burden associated with collecting additional data and concerns regarding the collection of data for demographic categories that could be considered sensitive, such as gender identity.

When asked about barriers to using their demographic data, most programs (90%) highlighted missing data or other problems with data quality as their primary concern ([Exhibit A-3](#)). Other barriers included insufficient staff to manage or analyze data (52%), insufficient IT infrastructure (48%), insufficient funding to support data analysis activities (45%), perception that available data were not aligned to needs of the program (31%), insufficient time to conduct analyses (28%), and difficulty formulating analytic goals or objectives from grantee data (21%). Another 10% of respondents identified other barriers, such as challenges related to the COVID-19 Public Health Emergency.

⁹ Per [Statistical Policy Directive 15 \(SPD15\)](#), "Wherever possible, race and/or ethnicity data should be collected through self-report, where the respondents directly provide their own race and/or ethnicity. In cases where self-report is not possible, data may be collected by proxy reporting, where a person knowledgeable of another's race and/or ethnicity responds on their behalf; by record matching, where existing records on an individual that contain their race and/or ethnicity are used to supply the information; or by observer identification, where an observer uses their best judgement of the most appropriate race and/or ethnicity categories in which to report an individual."

Program size, length of program, and lack of a comparison group were also cited as barriers to using demographic data. One respondent commented, “It’s too soon to tell. Our numbers are small, and we haven’t collected enough data to identify key findings.” suggesting a potential limitation in the usability of data for small programs or demonstrations with brief service distribution periods. Another respondent noted the challenges of understanding program data without also knowing the experiences of non-program beneficiaries, stating, “The data show that the program has been largely successful in attaining its goals, though this would be a more reliable finding if we were able to have a stronger comparison group.”

Exhibit A-3. Barriers to using HHS Demographic Data

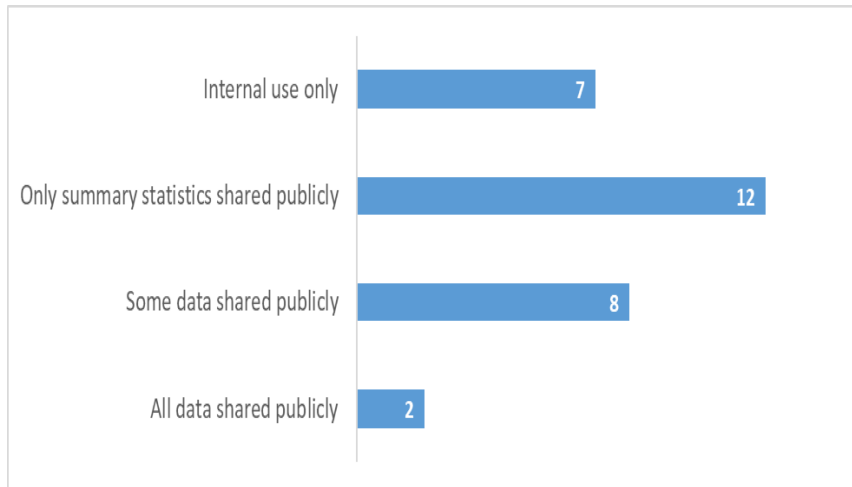


Because programs, agencies, and offices collected data that were program-specific, cross-program comparisons were often not feasible due to differences in the types of data collected or data collection methodologies.

Research Question 5: To what extent are demographic data available for external use by other HHS staff, the public, and other partners?

Most HHS demographic data were not publicly available ([Exhibit A-4](#)), with seven programs (24%) indicating their data were for internal use only and another twelve (41%) sharing that only some summary statistics were shared with the public. Eight programs (28%) said some data were available to the public, and two programs (7%) made all their data publicly available. Among the ten programs that shared at least some data with the public four programs’ data were accessible only upon

Exhibit A-4. Availability of HHS Demographic Data (n = 29)



request (e.g., shared via emailed link), three required special access via a restricted web portal, and three programs posted data online for download without special request or permission needed.

Research Question 6: How do larger program systems have to change to obtain data of sufficient quality for assessing disparities?

At the time of this study there was a lack of formal requirements for additional data collection beyond what was mandated by statutory and/or funding requirements. The lack of demographic data standardization resulted in wide variation in data quality within and across programs and made it difficult to identify specific systemic changes needed to identify and address service disparities. However, the survey results pointed to a need to revise federal data collection requirements.

Research Question 7: What are recommendations for improving the capacity of HHS demographic data systems to collect and improve demographic data elements, and promote their use for decision-making, program improvement, and research?

The RTI Project team identified three recommendations related to data validity, data completeness, and alignment of data items to program needs. These potential growth areas are detailed as follows.

Data Validation. The most valid and least error-prone method of demographic data collection is participant self-report.¹⁰ Where possible, programs reporting use of observation or other forms of proxy reporting should increase the use of self-report in their data collection processes in order to avoid potential bias. Programs using secondary data from other sources should confirm whether the data are collected through participant self-report to avoid over- or underestimating service utilization, program outcomes, and program impact.

Data Completeness. Several programs allowed participants to skip demographic questions that were not required to qualify for services. While it is important to minimize respondent burden, if these data were required programs would have the potential for a clearer picture of the populations they serve and the ability to better pinpoint service gaps. For example, requiring participants answer questions about language could help programs identify which languages are represented and needed for translation or interpretation services to ensure equitable service access. Insights gained from additional data could aid in decision making in a variety of areas including outreach and resource allocation.

Data Alignment. Standardization of data requirements, demographic variables, and data collection processes across HHS would make comparisons within and between HHS offices and agencies more feasible. It would also improve HHS's ability to create HHS-wide aggregate assessments of service reach and outcomes.

Limitations

This study was the first step in exploring HHS demographic data collection with an aim of improving program equity. The primary limitation of this study was the small sample size of 29 programs. Future studies might consider an expanded sampling frame to include more programs and an additional data collection mode that could facilitate more open-ended discussion such as interviews or focus groups.

¹⁰ Hammerton, G., & Munafò, M. R. (2021). Causal inference with observational data: the need for triangulation of evidence. *Psychological Medicine*, 51(4), 563–578. <https://doi.org/10.1017/S0033291720005127>

Mistrust and hesitation on the part of program staff and administration placed additional limitations on this study. Some program staff invited to participate indicated that they did not respond because they feared the survey would make their program “look lax” on data collection because they did not collect data across all 17 of the demographic categories. Other key program officials were guarded in how they approached the survey. For example, four different programs indicated that they would not respond to the survey until they received more information about who would have access to the survey data and what would be done with survey responses. Future efforts should take these concerns into account during study planning and recruitment.

To better understand how existing data requirements vary across programs, a more thorough evaluation of program-specific HHS data requirements and data collection guidelines (e.g., response options) is required. That activity was beyond the scope of this project, but it is clear from these findings that data collection requirements differ vastly across HHS programs. In addition, the fact that many programs are collecting demographic data beyond what is legislatively required suggests that statutory and/or funding requirements alone may be insufficient to address all program needs. This finding suggests that there are opportunities for HHS to expand existing data collection requirements within agencies and offices or across programs.

IV. CONCLUSION

This study found wide variation in the types of demographic data collected and data collection methodologies used by HHS programs both within and between HHS offices and agencies. Many programs collected limited participant demographic information, largely citing that additional data were not required or were not necessary for program administration. Existing data were used for a variety of internal purposes, including to understand the populations served and to identify needed program changes or service expansions. However, many program respondents also noted concerns about data quality and missingness that may reduce possible data uses. Finally, this study found that the majority of the demographic data collected by HHS programs are not share publicly. This means that use of those data was limited to program staff and that cross-program analyses were often not feasible.

These findings indicate that additional research is needed to better understand the nuances of program-specific data collection. Future efforts should aim to explore the data collection documents received from respondents in this study as well as a to better understand existing HHS data requirements by program. More primary data collection directly with federal employees to understand their program structures and experiences related to demographic data collection also would be beneficial, as well as conversations about reasons for not collecting some data, existing dataset accessibility, and the potential burden of collecting data across more domains. Additionally, another survey could reach more programs and could include more structural and social determinant domains (e.g., justice involvement) that were beyond the scope of this effort.

As HHS considers next steps, it is also worth noting that any new processes, data collection items, or analyses could benefit from the inclusion of program staff in the decision-making process. Engaging them in this way could help support buy-in, reduce hesitancy, and increase data quality. Likewise, inclusion of program participants in the design of future data collections or data modernization efforts also could help ensure that questions and response items are appropriate, representative, and meaningful to program recipients. Such efforts would support HHS in increasing service equity and outcomes for all populations.

Appendix A: Guiding Research Questions

RTI's survey of HHS programs was designed to address the following research questions:

1. What types of demographic data are HHS programs collecting about individuals enrolled in/served by HHS-funded programs?
2. Is the demographic data collected by HHS of sufficient quality to assess disparities in service utilization, program outcomes, and program impact?
3. To what extent are demographic data used to understand health and human-services related conditions, services, and outcomes of individuals and different communities?
4. What are the barriers to routinely collecting and using demographic data?
5. To what extent are demographic data available for external use by other HHS staff, the public, and other partners?
6. How do larger program systems have to change to obtain data of sufficient quality for assessing disparities?
7. What are recommendations for improving the capacity of HHS demographic data systems to collect and improve demographic data elements, and promote their use for decision-making, program improvement, and research?

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Office of the Assistant Secretary for Planning and Evaluation

200 Independence Avenue SW, Mailstop 447D
Washington, D.C. 20201

For more ASPE briefs and other publications, visit: aspe.hhs.gov/reports

ABOUT THE AUTHORS

Kevin C. Heslin, PhD (CDC)
Timothy Creedon, PhD (ASPE)
Ryan Mintz, MA (ASPE)
Helen Lamont, PhD (ASPE)
Maura R. Reilly, PhD, MPH (ASPE)

SUGGESTED CITATION

Heslin K., Creedon T., Mintz R., Lamont H., and Reilly M. The Availability, Quality, and Use of Administrative Data to Identify and Reduce Health and Human Services-Related Disparities: A Survey of HHS-Funded Programs and Interventions. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. April 2024.

COPYRIGHT INFORMATION

All material appearing in this report is in the public domain and may be reproduced or copied without permission. Citation as to source is appreciated.

DISCLOSURE

This communication was printed, published, or produced and disseminated at U.S. taxpayer expense.

ACKNOWLEDGMENTS

The project team thanks Sharon Arnold, Associate Deputy Assistant Secretary for Science and Data Policy (ASPE), for her key contributions to the initial concept and scope of work framing the project. Additionally, special thanks to the HHS Data Council, and in particular the HHS Data Disparities Workgroup and the Steering Group headed by Kevin Heslin (CDC), Rebecca Bunnell (CDC), Lisa Lang (NIH), Samantha Meklir (ONC), and Kristen Hudgins (ASPE).

This brief is based on a survey and report produced through a contract with RTI International.

To learn more about ASPE visit: aspe.hhs.gov/about